Knowledgeable Pretrained Language Models

刘 知 远
清华 大学

2020年12月19日
Natural Language Processing

• NLP aims to make computers understand languages
• The nature of NLP is structure prediction

NLP Is The Key of AI

NLP: The Key to Pass Turing Test and Realize AI
### Deep Learning: Data-Driven NLP

---

**Table and Diagram Details:**

- The diagram illustrates the process of translating a sentence from one language to another using deep learning.
- The table shows the probabilities of transitioning from one word to another in the source and target languages.
- The probabilities are represented by values between 0 and 1, indicating the likelihood of each transition.
- The diagram includes a sequence of words from the source language and their corresponding translations in the target language.

**Example:**

- The source sentence in the German language is: **Die proteste waren am wochende eskaliert**
- The target sentence in the English language is: **The protests escalated over the weekend**

---

**References:**

Language & Knowledge

- Knowledge enables people to understand language from superficial meanings to implicative meanings.

- We need to cure heatstroke.
- We need cooling devices.
- This summer is hot.
- The summer is like an oven.
- 这个夏天就像烤箱一样.
… we feel confident that more data and computation, in addition to recent advances in ML and deep learning, will lead to further substantial progress in NLP. However, the truly difficult problems of semantics, context, and knowledge will probably require new discoveries in linguistics and inference.
Pretrained Language Model as a Breakthrough in 2018

• Impressive progress of deep learning on unsupervised text corpora

- 2001: Neural language models
- 2008: Multi-task learning
- 2013: Word embeddings
- 2013: Neural networks for NLP
- 2014: Sequence-to-sequence models
- 2015: Attention
- 2015: Memory-based networks
- 2018: Pretrained language models

Sebastian Ruder http://ruder.io/a-review-of-the-recent-history-of-nlp/
What is Language Model

• Language models aims to predict the probability of a sequence as a natural sentence, or predict the probability of the next word given context
• Language models are a key to NLP and semantic representation of documents

她是中国人工智能领域的著名______
Challenge of Deep Learning in NLP

- Deep Learning has achieved the best performance in most NLP tasks
- Challenges: require large-scale supervised training
Pretrained Language Models

- Pre-trained Language Models (PLMs) can learn language patterns from large-scale un-labeled data, and improve the performance on downstream tasks by fine-tuning parameters.
Superior Performance on Language Understanding

![GLUE Benchmark Chart]

- CBOW: 58.6
- BiLSTM: 64.2
- BiLSTM+ELMO: 67.7
- Human: 87.1
- Roberta: 88.4
- ELECTRA: 89.4
- MT-DNN: 89.9
- T5: 90.3
- ALBERT: 90.6
Superior Performance on Language Generation

![Graph showing interactive SSA (%) vs. perplexity with data points for different models: Human (86%), Meena (79%), Meena (base) (72%), Cleverbot (56%), Mitsuku (56%), DialoGPT (48%), Xiaolce (31%).]
Contests of Pretrained Language Models

https://github.com/thunlp/PLMpapers
Contests of Pretrained Language Models

- **Encoder**
  - MLM
  - Permutated LM
  - Adversarial
  - **Decoder**
    - General
    - Dialogue
    - **Enc+Dec**
      - Joint
      - Separate
- **PLMs**
  - BERT, RoBERTa, SpanBERT, ERNIE
  - XLNet
  - ELECTRA
  - GPT, GPT-2, GPT-3, CTRL
  - Meena, Blender, DialoGPT
  - UniLM, T5
  - MASS, BART
Knowledgeable PLM

- External knowledge information can benefit language understanding, for low resource entities, and implicit background knowledge.

*Bob Dylan* wrote *Blowin' in the Wind* in 1962, and wrote *Chronicles: Volume One* in 2004.
How to Make PLMs Knowledgeable

• **Knowledgeable Input:** input augmentation as extra features

• **Knowledgeable Tasks:** knowledge-guided pre-training tasks

• **Knowledgeable Framework:** knowledge-guided neural architecture
Knowledgeable Input

- ERNIE: Enhanced Language Representation with Informative Entities
  - Lower layers for text, and higher layers for knowledge integration

Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, Qun Liu. ERNIE: Enhanced Language Representation with Informative Entities. ACL 2019.
Knowledgeable Input

- KnowBERT: Knowledge Enhanced Contextual Word Representations

Knowledgeable Input

• K-BERT: Directly add knowledge without further pre-training using knowledge layer.

Knowledgeable Tasks

- KEPLER: Joint learning of knowledge and language modeling
- Unify knowledge embedding and language representation into the same semantic space
Knowledgeable Tasks

• KEPLER: Joint learning of knowledge and language modeling

• Unify knowledge embedding and language representation into the same semantic space

Knowledgeable Framework

• K-Adapter: Inject multiple kinds of knowledge by training adapters independently on different tasks, support continual knowledge infusion

Knowledgeable Framework

- LM with mechanisms for selecting and copying facts from KG

Super Mario Land is a 1989 side-scrolling platform video game developed and published by Nintendo as a launch title for their Game Boy handheld game console.

## Resource: Chinese Pre-Trained Models (CPM)

<table>
<thead>
<tr>
<th>训练数据</th>
<th>模型大小</th>
<th>任务</th>
</tr>
</thead>
<tbody>
<tr>
<td>新闻</td>
<td>参数量</td>
<td>文本分类</td>
</tr>
<tr>
<td></td>
<td>109M, 334M, 2.6B</td>
<td></td>
</tr>
<tr>
<td>百科</td>
<td>层数</td>
<td>自然语言推理</td>
</tr>
<tr>
<td></td>
<td>12, 24, 32</td>
<td></td>
</tr>
<tr>
<td>对话</td>
<td>隐向量维度</td>
<td>阅读理解</td>
</tr>
<tr>
<td></td>
<td>768, 1,024, 2,560</td>
<td></td>
</tr>
<tr>
<td>网页</td>
<td>每层注意力数</td>
<td>完形填空</td>
</tr>
<tr>
<td></td>
<td>12, 16, 32</td>
<td></td>
</tr>
<tr>
<td>故事</td>
<td>注意力向量维度</td>
<td>对话生成</td>
</tr>
<tr>
<td></td>
<td>64, 64, 80</td>
<td></td>
</tr>
</tbody>
</table>

```python
print("输出：", tokenizer.decode(generates))
```

**输入:** 我们当中要数小明要厉害 -> 小明
文王昨天去了上海，之后又回到了武汉 -> 王文
他看到一个人，那人正是他的老师周治平 -> 周治平
我们都认为欧阳一凡的为人不错 ->

**输出:** 欧阳一凡

**输入:**
姚明的身高是211cm，是很多人心目中的偶像。 -> 姚明，身高，211cm
毛泽东是绍兴人，早年在长沙读书。 -> 毛泽东，出生地，绍兴
虽然周杰伦在欧洲办的婚礼，但是他就是土生土长的中国人。 -> 周杰伦，国籍，中国
小明出生于武汉，但是却不喜欢在武汉上学，长大后去了北京。 -> 小明，出生地，武汉
吴亦凡是很多人的偶像，但是他却是加拿大人，很多人失望。 -> 吴亦凡，国籍，加拿大
文王的生日在5月8号。这一天，大家都为他庆祝了生日。 -> 文王，生日，5月8号
《青花瓷》是周杰伦最得意的一首歌。 -> 周杰伦，作品，《青花瓷》
北京是中国的首都。 -> 中国，首都，北京

**输出:** 刘宕，生日，11月17号
Resource: Chinese Pre-Trained Models (CPM)

CPM-Generate
Chinese Pre-Trained Language Models (CPM-LM) Version-I
- Python
- MIT
- 54
- 595
- 9
- 0
Updated 2 days ago


CPM: A Large-scale Generative Chinese pre-trained Language Model
Authors: Zhengyan Zhang, Xu Han, Hao Zhou, Pei Ke, Yuxian Gu, Deming Ye, Yuja Qin, Yusheng Su, Haozhe Ji, Jian Guan, Fanchao Qi, Xiaozhi Wang, Yanan Zheng, Guoyang Zeng, Huanqi Cao, Shengqi Chen, Daixuan Li, Zhenbo Sun, Zhiyuan Liu, Minlie Huang, Wentao Han, Jie Tang, Juanzi Li, Xiaoyan Zhu, Maosong Sun

Abstract: ...as the training corpus of GPT-3 is primarily English, and the parameters are not publicly available. In this technical report, we release the Chinese Pre-trained Language Model (CPM) with generative pre-training on large-scale Chinese training data. To the best of our knowledge,...

Submitted 1 December, 2020; originally announced December 2020.
Open Source

- Packages for representation and acquisition of linguistic and world knowledge
- The projects obtain 40000+ stars on GitHub

https://github.com/thunlp
Books

**KG and DL**

**GNN**

**RL for NLP**

Open Access!
Outlook

- More methods to incorporate multiple knowledge into deep learning
Outlook

• More knowledge in future, concepts, commonsense, event, …
Knowledge-Guided NLP
Special Issue CFP on Pre-trained Language Models


Guest Editor: Zhiyuan Liu, Xipeng Qiu, Jie Tang

**AI Open Special Issue/Section on Pretrained Language Models**

**Call for Papers**

The release of ELMo, BERT, and GPT in 2018 indicates the success of pre-trained language models (PLMs), and the following years witness their great breakthrough on natural language understanding and generation. Many works have been done to explore more efficient and effective architectures for pre-training, to further improve pre-trained language models with cross-modal data, cross-lingual data, and structured knowledge, etc., or innovatively apply PLMs in various NLP-related tasks.

This special issue on Pretrained Language Models is devoted to gathering and presenting cutting-edge review, research, or applications of PLMs, providing a platform for researchers to share their recent observations and achievements in this active field. Specific topics for this special issue include but are not limited to:

- Novel architectures and algorithms of PLMs
- Generative PLMs
- Fine-tuning and adaptation of PLMs
- Multi-task and continual learning of PLMs
- Knowledge-guided PLMs
- Cross-lingual or multi-lingual PLMs
- Cross-modal PLMs
- Knowledge distillation and model compression of PLMs
- Analysis and probing of PLMs
- Applications of PLMs in various areas such as information retrieval, social computation, and recommendation
Thanks!

liuzy@tsinghua.edu.cn

http://nlp.csai.tsinghua.edu.cn/~lzy